CHAPTER 3

RESEARCH METHODOLOGY

3.1 INTRODUCTION

The data for the proposed research has been collected from web resource and glucose control projects of reputed universities. This Chapter attempts to give details about the data collection details and the research methodology of the proposed work modules carried out in the present research.

3.2 DATA COLLECTION

The proposed methods were developed with MATLAB (R2010a) and tested with three different data sets. The first data set have been obtained from the Glucosim, a web based diabetes simulator developed by Illinois Institute of Technology (Glucosim, 2006). For the given input conditions, this simulator gives the 24 hours blood glucose dynamics of a T1D Mellitus subject with 1 minute sampling frequency. Timing and dosage of Insulin and Carbohydrate (CHO) food intake, weight of a person and duration of exercise are the considered input parameters.

The second data set were gleaned from the glucose control project of the University of California, San Diego (UCSD, 2008). This glucose control project data base had BG values in 5, 10 and 20 minutes sampling interval for a one day period and some with two day period.

The third data set include the CGMS data for 25 days culled from a T2D subject (Rollins et al 2010). Actually this data set has the collection of other physiological parameters values also which were intended to be used in
modeling of Diabetes in free living conditions. However, the BG dynamics alone has been used for the current research work.

For clinical evaluation of the proposed methods, the BG data have been obtained from the diabetes subjects in some of the hospitals in Tamil Nadu, South India. The details of glucosim and clinical data collection are given in the following chapter.

Inexplicable fluctuations in BG dynamics are quiet common in diabetes mellitus. Since diabetes mellitus is a metabolic disorder, the secretion of insulin hormone and glucose absorption (the BG dynamics) will vary with individual to individual and in the same individual it will vary with emotions, food habits and exercise activities. The literatures (Scavini et al 2003, Juutilainen et al 2004, McGill et al 2013) show that blood glucose dynamics and control varies with gender, diabetes type and nature of work.

Scavini et al (2003) tested the hypothesis that diabetes and related risk factors are more common among female than male Zuni Indians. In the study carried out by McGill et al (2013), females in the analysis had smaller reductions in HbA1c and were less likely to reach glycemic goals despite higher insulin doses and more hypoglycemic events than males. Differences in gender responses to therapy should be considered when individualizing treatment for people with T2D. Adverse changes induced by T2D in some cardiovascular risk factors such as HDL cholesterol, triglycerides, LDL particle size, and blood pressure, have been found to be more pronounced in women than in men (Juutilainen et al 2004). And, it is possible that gender may alter the effect of some cardiovascular risk factors in diabetic subjects, leading to a stronger risk effect in women. Kilpatrick (2009) has given arguments for and against the role of glucose variability in the development of diabetes complications.
Hence, in the current research work, prediction of BG with three different approaches has been analysed with diabetes data sets of different categories such as male, female, type I or II, gestational diabetes, diabetes with hypertension, diabetes with sports activity, etc. To analyse the performance of the prediction models in different ranges, the proposed works have been tested in different blood glucose ranges like hypoglycemia, hyperglycemia and Normo glycemia. The performance of the three prediction models was investigated in terms of RMSE and Time lag values for PH of 30 and 60 minutes.

3.3 PROPOSED WORKS

The data from Continuous Glucose Monitoring devices are used by the diabetic clinicians to get information about the BG variations throughout the day, which facilitates them to arrive at optimal treatment decisions. The prediction of hypoglycemia is a clinically important task in the management of diabetes. Since hypoglycemia has dangerous effects such as, seizure and coma, it has to be predicted well in advance, and preventive measures should be adopted. An accurate prediction model is also needed for the implementation of an artificial pancreas.

However, this CGM data do have errors due to the time lag between venous blood and interstitial blood, improper calibration, sensor electronics, bio fouling and local inflammatory complications. The unwanted error components in the CGM signal should be removed to improve the quality of the signal so that the prediction approaches could produce better results. The proposed work has been formulated in two parts. First is the denoising of CGM sensor data with HFT and the second is the prediction of glucose concentration in blood plasma through three different approaches which are as follows, Prediction with time series ARIMA Model with regularized CGM data, Feature based Neural Network model (FNN) and prediction with customized ANFIS model.
In the Denoising module, a hybrid filtering technique has been applied. For reliable real time monitoring of BG, the filtering algorithm should account for:

1. Short term errors due to motion artifacts.
2. Random noise and other noise models.
3. Errors due to imperfect calibration.
4. Long term errors due to performance deterioration of sensor, biofouling, inflammatory complications etc.
5. Uncertainty in physiological parameters.

Filtering of short term errors and random errors for the linear case is simple. To account for errors due to imperfect calibration and sensor electronics, we go for nonlinear modelling, which could be achieved with Extended Kalman Filter (EKF). But to track the uncertainties in physiological parameters, some artificial intelligent modelling technique is needed. An artificial neural network, which does information processing as per the way of biological neurons would be a better choice.

The HFT comprises of a feed forward neural network trained with an EKF algorithm in back propagation. The overview of HFT is given in Figure 3.1

![Figure 3.1 Overview of Hybrid Filtering Technique](image)
HFT involves the development of a SS model with the dynamics of BG, IG and sensor gain deviation \( \phi \). This SS model with EKF algorithm was applied for the training of a feed forward neural network to denoise the errors in the CGM signal. The method was tested by applying the input CGM signal with WGN of different variance values. The time varying parameters like error covariance and the Kalman gain of the model were updated to meet the performance goal of minimum RMSE value. The detailed steps of HFT could be explained with a flow chart given in Figure 3.2.

In this work, the time varying nature of BG and sensor gain deviation parameter \( \phi \) are modeled with a AR model of order 2 and IG dynamics alone adopted taken as such from Rebrin et al (1999). The parameters were estimated with Least Squares (LS) method. The SS model, measurement model and the Kalman gain were computed by EKF algorithm. Then the Kalman gain parameter is used to customize the activation functions at hidden layer and output layer of the neural network. Then the NN is trained with the SS and measurement models so obtained until it reaches the stopping criteria of minimum RMSE.

To analyze the working of proposed denoising technique and prediction approaches in different diabetic subjects’ data and same subject with different blood glucose dynamics at different durations of the day, while training the NN, the CGM data is being added with WGN of different variance values in such a way to simulate inter individual variability and intra individual variability of blood glucose.

The performance of HFT has been compared with MA and Kalman filter in terms of RMSE in mg/dL, Time Lag in minutes and SRG. Analysis has been made with all the three data sets mentioned.
In the second part of work, three approaches have been framed, for prediction of glucose concentration in blood through ARIMA model with Tikhonov regularization, FNN and ANFIS. Since the physiological signals are strictly periodic, but rather fluctuate irregularly over time, analysis of physiological time series is complicated. Time series analysis is used to
identify systemic patterns like trends, seasonal patterns and methods for modeling and prediction. Preprocessing is necessary to extract the useful information underlying the time series, which is used for learning and to identify the model. Data regularization also improves the data quality and modelling.

After denoising the CGM sensor data signal with HFT, the proposed work applies Tikhonov regularization for smoothing ill posed CGM sensor data, and uses an ARIMA model to forecast the future glucose values. The regularization parameter was obtained with the standard L-curve procedure. The ARIMA model has been fixed with the Auto Correlation Function (ACF) and Partial ACF (PACF) plots and the parameter estimation done with LS method. Then the prediction performance of this model has been analyzed with an AR model, regularized AR model and ARIMA model with non regularized data to observe the improvement between AR and ARIMA models and the effect of regularization. RMSE and Time lag were the metrics used to assess the performance of proposed prediction models. The flow of works in CGM time series prediction with ARIMA modeling has been given in Figure 3.3.

The second prediction model is with AI. Since the interactions between the factors for glucose metabolism are complex, multidimensional, highly nonlinear, stochastically and time variant time series, the neural network model seems to be a more suitable predictor. It has been proved that, inclusion of past measurements increases the prediction accuracy of the neural network considerably. Glucose variability is an important factor in the development of diabetic complications. The parameters used nowadays to analyze the BG fluctuations are meant for retrospective analysis.
A prediction model incorporating the variability measure would be an optimum one. Hence, the proposed work FNN utilized the first four statistical moments of a time series along with approximate entropy as extracted features, for analyzing the fluctuations of BG dynamics and this multivariate model was applied for the training of an artificial neural network, for predicting the future glucose concentrations as shown in Figure 3.4.

In this part of work, a moving window of size fixed with false nearest neighborhood method has been applied. From each data window, the time varying features i.e., the first four standard moments of any time varying signal along with ApEn{Mean, Variance, Skewness, Kurtosis and Approximate Entropy} were extracted and applied for the training of FNN.
The development of a simple feed forward back propagation neural network with specialized transfer functions at hidden and output layer are presented in this section. The functions were designed to get activated, according to the learning parameter ‘$\alpha$’, which was derived from the features of the input data.

![Diagram of Feature Based Neural Network](image)

**Figure 3.4 Training phase of Feature based Neural Network (FNN)**

Here, the feature extraction has been used as the guidance in tracking the dynamics of nonlinear input. Thus, the neurons were made to generate the signal based on the derived feature, and not on instantaneous
values. Inspite of the predefined transfer / activation functions, custom made functions perform well as per the designed environment. In the proposed research work, two custom-made transfer functions were used, one at the hidden layer neurons and other at the output layer. This feature based NN has been trained to meet the performance measures of minimum RMSE and Time lag. The predicted output from this FNN has been compared with AR model and NN with LM algorithm.

![Diagram](Figure 3.5 Input selection procedure in ANFIS)
Vast metabolic biodiversity of diabetic population and limited knowledge on the complex human physiological process of glucose metabolism makes it difficult to achieve accuracy nearer to 100%. The third approach of prediction addresses the self tuning capacities of fuzzy systems and computational power of neural networks in predicting the BG concentration from continuous glucose monitoring sensor data time series. The aim of this study was to apply an ANFIS for the prediction of glucose concentration. The same set of five features \{Mean, Variance, Skewness, Kurtosis and ApEn\} were used here also, in which two have been selected based on their potential priority in representing the dynamics of the input data. The input selection procedure in ANFIS has been shown in Figure 3.5.

The ANFIS models have been trained with three input and two input combinations. The model with minimum training and testing errors was selected for the final prediction with fine tuning. The training phase has been customized with an LS estimation for consequent parameters and differential evolution based optimized estimation of premise parameters. The details of the forward pass and backward pass have been given in Figures 3.6 and 3.7 respectively. The performance of the customized ANFIS model has been compared the Reg. ARIMA and FNN models in terms of RMSE and Time lag. To confirm the efficiency of this custom ANFIS model, the performance is analysed with different types of data sets like T1D and T2D, male and female, gestational diabetes, diabetes with hypertension and diabetes with sports activity and data ranges like hypo glycemia, hyper glycemia and normo glycemic ranges. On comparing the results of the three proposed prediction models, the performance of the customized ANFIS model is better in all types of data sets and all types of data ranges.
Layer 1
- Initialize the premise parameters \( C_i \) with mean and \( \sigma_i \) with variance of input data set.
- Compute the membership values \( (\mu_{H_i}(x), \mu_{G_i}(y)) \)

Layer 2
Product of membership values
i.e. the firing strength of rules, \( W_i \)

Layer 3
Normalize the firing strengths \( W_i \)

Layer 4
- Find the Consequent parameters \( p_i, q_i, r_i \) through Least Squares method.
- Calculate the Rule Outputs \( f_i \)

Layer 5
Calculate the predicted output from ANFIS model in layer 5, \( F_k \)

Figure 3.6 Forward pass in ANFIS training
3.4 SUMMARY

In this chapter, the details of research methodology were presented along with an insight to data sets used for testing the performance of the prediction models. The next chapter elaborates the details of data collection used for the current research works.